(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072627 A2

(51) International Patent Classification[7]: C07K 14/47, A61K 39/00, A61P 35/00, G01N 33/68, G06F 19/00

(21) International Application Number: PCT/EP02/02666

(22) International Filing Date: 11 March 2002 (11.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/274,250	9 March 2001 (09.03.2001)	US
60/290,353	14 May 2001 (14.05.2001)	US
60/291,610	18 May 2001 (18.05.2001)	US

(71) Applicant *(for all designated States except US)*: **CALLISTOGEN AG** [DE/DE]; Neuendorfstrasse 24b, 16761 Hennigsdorf (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WREDE, Paul** [DE/DE]; Reichensteiner Weg 7, 14195 Berlin (DE). **WALDEN, Peter** [DE/DE]; Rykestrasse 4, 10405 Berlin (DE). **EICHLER-MERTENS, Mathias** [DE/DE]; Gripsstrasse 16, 10119 Berlin (DE). **FILTER, Matthias** [DE/DE]; Seestrasse 26, 15518 Petersdorf (DE).

(74) Agents: **WEICKMANN, Franz, Albert** et al.; Weickmann & Weickmann, Postfach 860 820, 81635 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** INTRODUCTION OF ANTI-TUMOR T LYMPHOCYTES IN HUMAN USING PEPTIDE EPITOPES FOUND BY COMPUTER BASED ALGORITHMS FOR VACCINATION

(57) **Abstract:** This invention relates to a method for providing, identifying or/and optimizing peptides which induce cytotoxic T-lymphocytes and to the uses of the thus obtained peptides, in particular, for vaccination.

WO 02/072627 A2

Induction of anti-tumor cytotoxic T lymphocytes in humans using peptide epitopes found by computer based algorithms for vaccination

Description

This invention relates to a method for providing, identifying or/and optimizing peptides which induce cytotoxic T-lymphocytes and to the uses of the thus obtained peptides, in particular, for vaccination.

In particular, this invention relates to a method for predicting and optimizing peptides and peptidomimetics, based on the application of pattern recognition technologies such as, for example, artificial neural networks, in combination with a selection for the highest degree of conservation, in particular, phylogenetic conservation and optimization through amino acid exchange at the anchor positions of the MHC-binding peptides, and the use of the identified amino acid sequences in a peptide pool, e.g. together with additional helper antigens as co-stimulators for vaccination.

The present invention further relates to compositions and methods for the treatment of cancer and the treatment or prevention of viral infections. The invention, in particular, provides peptides based on a 9 residue epitope derived from tumor-associated or viral antigens. The peptides induce cytotoxic T cells that destroy tumor cells and virus-infected cell.

Further, this invention relates to computer-assisted analysis of biological molecules, particularly of biologically active peptides and peptide mimetics, and the prediction of their biological and pharmacological potencies.

Vaccines against tumors or viruses are based on specific antigens, in particular, on weakly immunogen-specific antigens, admixed to adjuvants in order to elicit, restore or augment immune responses against tumor cells, e.g. residual or metastatic tumor cells, or virus-infected cells. Cellular cytotoxicity is considered to play a major role in the elimination of tumor cells or virus-infected cells. Activation of cellular cytoxicity within an organism requires at least three synergistic signals: Epitopes derived from tumor-specific antigens presented by MHC class I molecules (HLA restriction), co-stimulatory signals provided by cell surface molecules of antigen-presenting cells (APCs), e.g. B-7.1 and B-7.2, and differentiation and propagation signals of cytokines.

To activate cellular cytotoxicity it is therefore of great interest to find and/or provide pertinent HLA-restricted epitopes, especially also in view of the widespread occurrence of cancer and viral diseases. Therefore, it was an object of the invention to provide peptides which induce cytotoxic T-lymphocytes.

According to the invention this object is achieved by a method for providing, identifying or/and optimizing peptides which induce cytotoxic T-lymphocytes, comprising the steps:

(a) selecting one or more antigenic proteins,
(b) selecting conserved regions within the protein sequence of the one or more antigenic proteins, and
(c) identifying CD8+ T-cell epitopes within the protein sequence of the one or more antigenic proteins, preferably within the phylogenetically conserved regions.

According to the method of the invention one or more antigenic proteins are selected in a first step. In particular, relevant antigenic proteins for various cancers or viruses are taken. The selection can be performed, for

example, by the man skilled in the art referring to literature or references describing antigenic proteins associated with cancers and viruses.

In a second step, conserved regions within the protein sequence of one or more antigenic proteins are determined. The determination of conserved regions can be effected, for example, by comparison with other proteins, e.g. proteins stored in a database. In step (b) according to the invention conserved regions, i.e. regions which are subject only to minor changes during evolution, are determined. The selection of conserved regions, in particular, has the advantage that a high response rate is achieved in subsequent use of the peptides for inducing cytotoxic T-lymphocytes, and high effectiveness against the cancer cells and viruses to be attacked. In contrast to highly variable regions, conserved regions change only slightly and, thus, represent an excellent target for combatting cancer cells or viruses. It is especially preferable to select phylogenetically conserved regions within the protein sequences of the one or more antigenic protein.

In a further step according to the invention CD8+ T-cell epitopes are identified within the protein sequence of the one or more antigenic proteins and preferably within the conserved regions, in particular, within the phylogenetically conserved regions. Determination of CD8+ T-cell epitopes can be effected by means of pattern recognition technologies and, especially by using an artificial neural network (ANN). Artificial intelligence and pattern recognition methods have been proven to be powerful tools in the bioinformatics field. For example, an artificial neural network (ANN) has been successfully applied to predict mitochondrial precursor cleavage sites (G.Schneider, P.Wrede, J.Mol.Evol.36, 586 (1993) and membrane-spanning amin acid sequences (R.Lohmann, G.Schneider, D.Behrens and P.Wrede, Protein Science 3, 1597 (1994); M.Milik and J.Skolnick, in: "Proceedings of Fourth Annual Conference on Evolutionary Programming", MIT Press, La Jolla (1995)).

However, the identification of CD8+ T-cell epitopes or the prediction of MHC-I binding can be done by any technology available to the man skilled in the art. In particular, pattern recognition technologies can be applied. Preferably, however, an artificial neural network is used, since an ANN allows for prediction of MHC-I binding peptides with high accuracy. Particularly preferred an ANN is used which has been trained with an evolutionary algorithm.

In a preferred and advantageous embodiment, the method according to the invention further comprises the step:

(d) optimizing the identified CD8+ T-cell epitopes by exchanging one or more amino acids.

Preferably, the amino acids are exchanged in the anchor positions of the epitopes, in particular, in the anchor residues of the MHC-1 binding peptides. Particularly preferred, said optimizing step is performed prior to the step of identifying CD8+ T-cell epitopes. According to the invention modified epitopes, too, are thus tested for their binding efficacy, as a result of which new effective peptides can be found.

Optimization of the CD8+ T-cell epitopes is preferably effected by exchanging the amino acid present by another amino acid at one or more positions of the peptides. Said exchange can be effected randomly and at arbitrary positions. It is preferred, however, to first determine anchor positions and then exchange the amino acids present at said anchor positions. Preferably amino acids are taken in exchange which are known to increase binding to MHC-I at these anchor positions.

By means of the method of the invention, in particular, peptides having a length of from 4-30, more preferably from 5-20, still more preferably of at least 6, at least 7, at least 8 or at least 9 amino acids, and up to 15, 14, 13, 12, 11 or 10 amino acids are obtained. It is particularly preferred to

apply the invention to peptides having a length of 8, 9 or 10 amino acids, especially 9 amino acids.

The term peptide as used herein also includes peptide mimetics which contain one or more non-naturally occurring amino acid, e.g. homoarginine, ornithine, etc.

Selection of suitable peptides which induce cytotoxic T-lymphocytes can be effected by means of the above-described procedural steps, in particular, by selecting the respective best candidates of each procedural step, e.g. the best 50%, the best 30% or the best 10%. In addition, it is possible to incorporate filtering steps, by means of which particular peptides are selected and picked out as preferred or disposed of.

According to the invention the predicted identified or optimized epitope peptides can be verified by in vitro or in vivo tests, especially by in vitro tests.

The peptides obtained according to the invention, finally, can be used as pharmaceuticals, especially as a vaccine. In particular, tumors and virus infections can be treated or prevented successfully by means of the peptides obtained according to the invention.

Therefore, the invention further relates to a pharmaceutical composition comprising one or more peptides obtainable by the method described above. This pharmaceutical composition can comprise further adjuvants, co-factors and/or co-stimulating agents, e.g. recall antigens as adjuvants for CD4* T-cell stimulation and for induction of co-stimulation for peptide and disease-specific CD8* cytotoxic T-cells. Particularly preferred, the pharmaceutically composition is a vaccine, in particular, a vaccine for the treatment and/or prevention of cancer or viral infections. The

pharmaceutical composition can be in any suitable administration form, with intracutaneous and parenteral administration being preferred.

An important and most preferred aspect of the invention is the combination of methods to identify peptides and the subsequent use of the peptides found as pharmaceutical composition, in particular, for vaccination. Therefore, a most preferred embodiment of the invention is a method for providing a pharmaceutical composition for the induction of cytotoxic T-lymphocytes cmoprising:

(a) providing one or more peptides which induce cytotoxic T-lymphocytes according to the method described above, and
(b) using the one or more peptides for the manufacture of a pharmaceutical composition.

The invention allows, in a unique manner, to combine these two steps. In particular, the invention allows to actually provide pharmaceuticals, starting out from computer-based predictions.

The invention further relates to the peptides discovered by means of the inventive method, in particular, as shown in Tables 1, 2, 3 and 4 below, as well as to pharmaceutical compositions containing one or more of these peptides or other peptides discovered by means of the method of the invention, in particular, at least 2, at least 3, at least 4, at least 5, at least 10 or at least 20 and up to 100, preferably up to 90, up to 80, up to 70, up to 60 or up to 50 of such peptides.

Further peptide sequences of the invention are as shown in the following. In these sequences the amino acid at positions 2, 6 or/and 9 each independently can be replaced by V, L, I or/and M.

Table 1

::::::::::::::		
catd_human		
::::::::::::::		
1.000000	YLSQDTVSV	150-158
0.999313	KLVDQNIFS	222-230
1.000000	LVDQNIFSF	223-231
0.997769	DQNIFSFYL	225-233
0.989067	VTRKAYWQV	264-272
0.999877	QVHLDQVEV	271-279
0.999934	HLDQVEVAS	273-281
::::::::::::::		
creb_human		
::::::::::::::		
0.916947	ILNDLSSDA	137-145
0.998294	TTILQYAQT	219-227
0.989879	TILQYAQTT	220-228
0.997264	DVQTYQIRT	248-256
0.999142	AARKREVRL	282-290
0.999527	AVLENQNKT	316-324
0.999975	VLENQNKTL	317-325
0.999923	TLIEELKAL	324-332
::::::::::::::		
ctag_human		
::::::::::::::		
0.999339	ELARRSLAQ	103-111
0.999982	VLLKEFTVS	121-129
0.999974	NILTIRLTA	131-139
0.999991	ILTIRLTAA	132-140
0.998567	TIRLTAADH	134-142
0.999960	AADHRQLQL	139-147
::::::::::::::		
erb2_human		
::::::::::::::		
0.999709	SFLQDIQEV	72-80
0.999802	LIAHNQVRQ	85-93
0.999996	QLFEDNYAL	106-114
0.999996	LFEDNYALA	107-115
0.974558	QLRSLTEIL	141-149
0.998018	TILWKDIFH	166-174
0.611272	ILWKDIFHK	167-175
0.999929	DIFHKNNQL	171-179
0.996131	KNNQLALTL	175-183
0.947400	NNQLALTLI	176-184
0.999993	QLALTLIDT	178-186

0.999957	LIDTNRSRA	183-191
0.999927	ALVTYNTDT	270-278
0.999967	LVTYNTDTF	271-279
0.998428	HLREVRAVT	349-357
0.834140	AVTSANIQE	355-363
0.999736	VTSANIQEF	356-364
0.999164	QVFETLEEI	398-406
0.922432	VFETLEEIT	399-407
0.999993	SVFQNLQVI	423-431
0.999985	ALIHHNTHL	466-474
0.704970	LTSIISAVV	651-659
0.999930	LIKRRQQKI	674-682
0.990487	RLLQETELV	689-697
0.999361	ETELRKVKV	717-725
0.998516	AIKVLRENT	751-759
0.891765	LTSTVQLVT	790-798
0.999969	STVQLVTQL	792-800
0.999802	YLEDVRLVH	835-843
0.999964	RLVHRDLAA	840-848
0.999992	DLAARNVLV	845-853
0.999993	LLDIDETEY	869-877
0.999885	DIDETEYHA	871-879
0.999956	SILRRRFTH	893-901
0.997836	ILRRRFTHQ	894-902
0.999221	RFTHQSDVW	898-906
0.884728	THQSDVWSY	900-908
0.999962	RFRELVSEF	968-976
0.999439	FVVIQNEDL	986-994
0.990193	DLVDAEEYL	1016-1024
0.999995	LVDAEEYLV	1017-1025

::::::::::::::
gp100_human
::::::::::::::

0.986531	QVIWVNNTI	101-109
1.000000	VIWVNNTII	102-110
0.991766	SWSQKRSFV	142-150
0.999969	SFVYVWKTW	148-156
0.999897	SVSVSQLRA	216-224
0.999997	YLAEADLSY	250-258
0.990549	VTAQVVLQA	286-294
0.999610	TTAAQVTTT	413-421
0.996788	AAQVTTTEW	415-423
0.983375	VTTTEWVET	418-426
0.999911	SFSVTLDIV	482-490
0.999597	NVSLADTNS	568-576
0.999882	SLADTNSLA	570-578
0.994679	LADTNSLAV	571-579
0.998051	HSSSHWLRL	632-640

::::::::::::::
mage1_human
::::::::::::::

0.999970	ALEAQQEAL	15-23
0.999915	ILESLFRAV	93-101
1.000000	VITKKVADL	101-109
0.927034	ASESLQLVF	147-155
0.978865	KLLTQDLVQ	237-245
0.999998	LVQEKYLEY	243-251
0.996432	LAETSYVKV	271-279
0.999888	YVKVLEYVI	276-284
0.999949	KVLEYVIKV	278-286
0.988293	YVIKVSARV	282-290
0.999463	KVSARVRFF	285-293

::::::::::::::

mage2_human

::::::::::::::

0.951325	ATEEQQTAS	32-40
0.951615	QTASSSSTL	37-45
0.938975	SFSTTINYT	70-78
0.999958	STTINYTLW	72-80
0.999946	TINYTLWRQ	74-82
0.999920	DLESEFQAA	100-108
1.000000	LVHFLLLKY	116-124
0.952279	HFLLLKYRA	118-126
0.999993	VIFSKASEY	149-157
0.999998	LVQENYLEY	250-258
0.999988	LIETSYVKV	278-286
0.989867	YVKVLHHTL	283-291
0.999986	KVLHHTLKI	285-293

::::::::::::::

mage3_human

::::::::::::::

0.962244	AASSSSTLV	38-46
0.999920	DLESEFQAA	100-108
1.000000	ALSRKVAEL	108-116
0.999951	KVAELVHFL	112-120
0.994726	VAELVHFLL	113-121
1.000000	LVHFLLLKY	116-124
0.952279	HFLLLKYRA	118-126
0.999930	VIFSKASSS	149-157
1.000000	IFSKASSSL	150-158
0.989063	ASSSLQLVF	154-162
1.000000	KIWEELSVL	220-228
0.999632	KLLTQHFVQ	244-252
0.999378	LLTQHFVQE	245-253
0.999996	FVQENYLEY	250-258
0.999978	LVETSYVKV	278-286

::::::::::::::

mage4_human

::::::::::::::

0.998536	TTEEQEAAV	32-40
0.999985	ALSNKVDEL	109-117
0.997846	KVDELAHFL	113-121
0.999083	HFLLRKYRA	119-127
0.982666	KLLTQDWVQ	245-253
0.991132	LLTQDWVQE	246-254
0.999989	WVQENYLEY	251-259
0.996432	LAETSYVKV	279-287
0.999961	KVLEHVVRV	286-294
0.998258	HVVRVNARV	290-298

::::::::::::::

mage5_human

::::::::::::::

0.999973	AIDFTLWRQ	74-82
0.999964	DFTLWRQSI	76-84
1.000000	ALSKKVADL	108-116
0.999983	KVADLIHFL	112-120
0.997569	VADLIHFLL	113-121
1.000000	LIHFLLLKY	116-124

::::::::::::::

mage6_human

::::::::::::::

0.962244	AASSSSTLV	38-46
0.999920	DLESEFQAA	100-108
1.000000	ALSRKVAKL	108-116
0.999978	KVAKLVHFL	112-120
1.000000	LVHFLLLKY	116-124

0.952279	HFLLLKYRA	118-126
0.999520	VIFSKASDS	149-157
0.998461	IFSKASDSL	150-158
0.975947	ASDSLQLVF	154-162
1.000000	KIWEELSVL	220-228
0.999880	KLLTQYFVQ	244-252
0.976478	LLTQYFVQE	245-253
0.999996	FVQENYLEY	250-258
0.999988	LIETSYVKV	278-286

::::::::::::::
mage8_human
::::::::::::::

0.998794	AASSSSTLI	38-46
0.999999	SLTVTDSTL	71-79
1.000000	ALDEKVAEL	111-119
0.997295	VAELVRFLL	116-124
0.999966	RFLLRKYQI	121-129
0.998206	SVIKNYKNH	141-149
0.999915	VIKNYKNHF	142-150

::::::::::::::
mage9_human
::::::::::::::

0.999999	SISVYYTLW	68-76
0.999512	SVYYTLWSQ	70-78
1.000000	ALKLKVAEL	107-115
0.999951	KVAELVHFL	111-119
0.994726	VAELVHFLL	112-120
0.998422	LVHFLLHKY	115-123
0.999920	HFLLHKYRV	117-125
0.999636	SVIKNYKRY	137-145
0.999991	EVIWEALSV	218-226
0.982666	KLLTQDWVQ	243-251
0.991132	LLTQDWVQE	244-252
0.999989	WVQENYLEY	249-257
0.902355	TSYEKVINY	280-288

::::::::::::::
mageA_human
::::::::::::::

0.997990	AVEEDASSS	33-41
0.999999	EIDEKVTDL	133-141
0.999324	KVTDLVQFL	137-145
0.999623	VTDLVQFLL	138-146
1.000000	LVQFLLFKY	141-149
0.999762	ILESVIKNY	160-168
0.997688	SVIKNYEDH	163-171
0.997563	VIKNYEDHF	164-172
0.982666	KLLTQDWVQ	269-277
0.991132	LLTQDWVQE	270-278
0.999989	WVQENYLEY	275-283
0.999984	SLLKFLAKV	310-318

::::::::::::::
mageB_human
::::::::::::::

0.996542	QAEEQEAAF	32-40
0.999977	AFFSSTLNV	39-47
1.000000	ILHDKIIDL	111-119
0.999993	KIIDLVHLL	115-123
1.000000	IIDLVHLLL	116-124
0.999952	HLLLRKYRV	121-129
0.999894	SVIKNYEDY	141-149
0.999975	YVLVTSLNL	179-187
0.989319	VLVTSLNLS	180-188
0.999988	LVTSLNLSY	181-189

0.986562	RLLTQNWVQ	247-255
0.999551	LLTQNWVQE	248-256
0.999999	WVQEKYLVY	253-261
0.999671	KVLEYIANA	288-296

::::::::::::::

mageC_human

::::::::::::::

0.941417	ETASSSSTL	37-45
0.999983	TINYTLWSQ	74-82
0.999919	DLETSFQVA	100-108
1.000000	LVHFLLLKY	116-124
0.952279	HFLLLKYRA	118-126
0.999995	SVIRNFQDF	138-146
0.995788	VIRNFQDFF	139-147
0.999993	VIFSKASEY	149-157
1.000000	KIWEELSVL	220-228
0.978865	KLLTQDLVQ	244-252
0.999998	LVQENYLEY	250-258
0.999978	LVETSYVKV	278-286
0.984092	YVKVLHHLL	283-291
0.999983	KVLHHLLKI	285-293

::::::::::::::

mdm2_human

::::::::::::::

0.999224	LLLKLLKSV	33-41	30
0.911561	SVKEHRKIY	92-100	6
0.998606	VVVNQQESS	108-116	7
0.982799	STSSRRRAI	157-165	15
0.999689	AISETEENS	164-172	12
0.912657	RHKSDSISL	183-191	11
0.999287	SISLSFDES	188-196	12
0.999913	SLSFDESLA	190-198	15
0.998285	SVSDQFSVE	240-248	7
0.999960	SVEFEVESL	246-254	20
0.999996	SLDSEDYSL	253-261	25
0.999888	IIYSSQEDV	403-411	21

::::::::::::::

mif_human

::::::::::::::

0.999946	FLSELTQQL	18-26
0.999957	ELTQQLAQA	21-29
0.942786	LLAERLRIS	82-90

::::::::::::::

p53_human

::::::::::::::

0.999372	ETFSDLWKL	17-25
0.999971	TFSDLWKLL	18-26
0.916063	NTFRHSVVV	210-218
0.999934	ALELKDAQA	347-355

::::::::::::::

tyr2_human

::::::::::::::

1.000000	VIRQNIHSL	125-133
0.999786	ALDLAKKRV	144-152
0.999993	SVYDFFVWL	180-188
0.999999	FVWLHYYSV	185-193
0.999976	FVTWHRYHL	216-224
0.994160	VTWHRYHLL	217-225
0.999975	TLISRNSRF	271-279
0.977456	SRNSRFSSW	274-282
0.999984	SLDDYNHLV	288-296
1.000000	FFQNSTFSF	339-347
1.000000	SLHNLVHSF	367-375

0.999237	IFVVLHSFT	391-399
0.999957	VVLHSFTDA	393-401
0.995790	VLHSFTDAI	394-402
0.955026	VTNEELFLT	439-447
0.990634	ELFLTSDQL	443-451
0.955259	HLSSKRYTE	509-517
::::::::::::::		
tyro_human		
::::::::::::::		
0.999956	LLWSFQTSA	9-17
0.999999	RLLVRRNIF	116-124
1.000000	LVRRNIFDL	118-126
0.999147	KFFAYLTLA	133-141
0.999998	YLTLAKHTI	137-145
0.999987	TLAKHTISS	139-147
0.987547	AKHTISSDY	141-149
0.999999	DINIYDLFV	169-177
0.995893	FLLRWEQEI	214-222
0.999994	SFFSSWQIV	267-275
0.999999	IFLLHHAFV	385-393
0.999924	LLHHAFVDS	387-395
0.997281	AFVDSIFEQ	391-399
0.999962	FVDSIFEQW	392-400
0.998227	SIFEQWLRR	395-403
0.927015	YLEQASRIW	467-475
0.979646	ASRIWSWLL	471-479

Further peptide sequences of the invention are as shown in the following. In these sequences the amino acids at positions 2, 6 or/and 9 each independently can be replaced by V, L, I or/and M.

Table 2

Pos.	Sequence	modification	Identity-scores		Comments
BCL2_HUMAN					
154	RIVAFFEFI	G -> I Pos 9	187	229	
137	RFATVVEEL		127	188	
188	YLNRHLHTW		124	188	
CCEM_HUMAN					
25	RLLLTASLL		203	237	
26	LLLTASLLT		209	237	
27	LLTASLLTF		210	236	
28	LTASLLTFW		210	236	
108	IIYSNASLL	P -> S Pos 4	183	229	
CD19_HUMAN					
427	EFYENDSNL		35	44	
326	VLRRKRKRI	M -> I Pos 9	35	44	
302	AVTLAYLIF		31	41	
287	VLWHWLLRT		30	41	
CGD1_HUMAN					
63	SLRKIVATW	M -> L Pos 2	431	709	
92	YLDRFLSLI	E -> I Pos 9	491	656	
152	LVNKLKWNL		320	630	
CTAG_HUMAN					
129	VLLKEFTVS		24	24	6,069,233
J Exp Med 2000 Feb 21;191(4):625-30: 15 AS epitope for MHC-II					
ERB2_HUMAN (Her-2)					
827	RLVHRDLAA		687	802	
6,037,135: 10 AA (RLVHRDLAA R); Seq-Id 288 for HLA-A3.2					
6,075,122: identical sequence patented Seq ID 18					
832	DLAARNVLV	I/L at pos. 9 often	789	860	
6,075,122: identical sequence patented Seq ID 9					
885	RFTHQSDVW		611	817	
MUC1_HUMAN					
1049	SFFFLSFHI		42	42	
1139	RYNLTISDV		39	39	
1061	QFNSSLEDI	P -> I Pos 9	44	44	
TRSR_HUMAN					
271	TFAEKVANA		219	287	
413	VIAQRDAWI	G -> I Pos 2 + 9	232	312	
455	SIIFASWSA		251	332	
489	YINLDKAVL		222	293	
TYR2_HUMAN					
188	SVYDFFVWL		111	147	Cancer Res. 1998;58(21):4895
193	FVWLHYYSV		124	148	
6,083,703: 10 AA peptide Seq-Id: 17; no activity seen in test					
6,132,980: s.o.					
224	FVTWHRYHL		128	168	
282	SRNSRFSSW		111	146	
351	STFSFRNAL		106	144	

CATD_HUMAN

106	TISSNLWVI	G, P -> I Pos 2,9	725	811
272	VTRKAYWQV		354	543
404	VFDRDNNRV		456	602

Immunogenetics 1996;43(6):392-7 18-mer as ligand

PM17_HUMAN

258	YLAEADLSY		47	56
294	VTAQVVLQA		45	59
576	NVSLADTNS		48	56

CREB_HUMAN

141	SYRKILNDL		115	124
325	VLENQNKTL		104	104

P53_HUMAN

25	ETFSDLWKL		197	216
218	NTFRHSVVV		263	281
257	RIILTIITL	P -> I Pos 2	295	303
355	ALELKDAQA		195	223

MIF_HUMAN

26	FLSELTQQL		73	103

MAG1_HUMAN

117	LVHFLLLKY	G -> H Pos 3 == MAG2	126	149

6,037,135: seq-ID 1205; HLA-3 and 11 binding; no CTL response
J Immunol 1999 Sep 1;163(5):2928-36: 14-mer with T-cell response

136	ILESVIKNY	M -> I Pos 1 == MAGA	111	138
129	ELVTKAEIL	M -> I Pos 8 == MAGA	130	150
		P -> L Pos 2 == MAG4		
155	ASESLQLVF		112	135
245	KLLTQDLVQ		117	130
251	LVQENYLEY	K -> N Pos 5 == MAG2	119	137
279	LIETSYVKV	A -> I Pos 2 == MAG2	103	130

6,147,187: Ser-ID 11; HLA-2.1 -> clearly claimed

Further peptide sequences of the invention are as shown in the following. In these sequences the amino acids at positions 2, 6 or/and 9 each independently can be replaced by V, L, I or/and M.

Table 3

Protein (Swiss-Prot-ID)	Peptide sequence	Position in the protein	Note	
VGR3_HUMAN	DLAARNILL	1037-1045		
	TTQSDVWSF	1092-1100		
	VLLWEIFSL	1102-1110		
VEGF_HUMAN	TLVDIFQEY	57-65		
CD34_HUMAN	ILDFTEQDV	272-280		
	TLIALVTSI	290-298	at pos9: G ->I	
	TIQATSRNI	364-372	at pos9: G ->I	
ETS1_HUMAN	QLWQFLLEL	336-344		
PEC1_HUMAN	VIVNNKEKT	111-119		
	IIIQKDKAI	270-278		
	SIVVNITEL	316-324		
MDM2_HUMAN	SVKEHRKIY	92-100		
MM01_HUMAN	HLTYRIENY	113-121		
	AFQLWSNVT	137-145		
	LHRVAAHEL	212-220		

Further peptide sequences of the invention are as shown in the following. In these sequences the amino acid at positions 2, 6 or/and 9 each independently can be replaced by V, L, I or/and M.

Table 4

position	sequence	Filter	conservation score	conservation score	ANN score
rrp2					
442-450	RRNYFTAEV	1	253	272	0.732
659-667	SAESRKLLL	1	262	271	0.873
510-518	HLRNDTDVV	1	264	268	0.753
701-709	LLNASWFNS	1	247	268	0.973
417-425	LTDSIWIEL	1	230	268	0.948
420-428	SIWIELDEI	1	232	268	0.980
638-646	RTLLAKSVF	1	237	267	0.900
rrp3					
546-554	LVNTYQWII	1	306	327	0.982
736-744	KRKRNSSIL	1	274	326	0.860
496-504	VSIDRFLRV	1	302	325	0.502
226-234	SVYIEVLHL	1	303	325	0.992
19-27	ILTKTTVDH	1	306	324	0.965
544-552	SVLVNTYQW	1	304	324	0.992
hema					
51-59	EVTNATELV	1	679	825	0.618
385-393	STQAAIDQI	1	767	818	0.798
435-443	DLWSYNAEL	1	720	817	0.985
463-471	LFEKTRRQL	1	668	815	0.925
245-253	RISIYWTIV	1	656	815	0.969
447-455	LENQHTIDL	1	715	810	0.933
382-390	DLKSTQAAI	1	755	800	0.837
380-388	AADLKSTQA	1	748	800	0.741
vmt1					
153-161	QIADSQHRS	1	155	179	0.738
180-188	VLASTTAKA	1	162	177	0.990
232-240	DLLENLQAY	1	155	171	0.953
102-110	KLKREITFH	1	149	171	0.555
vmt2					
35-43	ILHLILWIL	1	9	143	0.998
83-91	AVDADDSHF	1	129	142	0.989
39-47	ILWILDHLF	1	24	142	0.973
nram					
217-225	SWSKNILRT	1	380	462	0.995
438-446	WTSNSIVVF	1	309	436	0.967
437-445	WWTSNSIVV	1	305	416	0.895
435-443	RVWWTSNSI	1	287	406	0.961
389-397	KLQINRQVI	1	245	356	0.984
222-230	ILRTQESEC	0	473	492	0.993
02 - 10	NPNQKIITI	0	416	429	0.949
vnb					
28-36	SFTVILTVF	1	94	98	0.998
03-11	NATFNYTNV	1	96	96	0.913

Particularly preferred are peptides VTAQVVLQA, VLAQVVLQL, LVHFLLLKY, LLHFLLLKL, FVWLHYYSV or FLWLHYYSL, which showed particularly high activity in step (b) as well as variants generated by AA exchange at position 2, 6 and/or 9, e.g. by V, L, I or M.

The invention further relates to the use of the peptides found by the method of the invention for the production of a pharmaceutical for the induction of cytotoxic T-lymphocytes, in particular, for the prevention, treatment or diagnosis of cancer or viral infections.

The invention and the individual procedural steps will be explained in detail below.

HLA-restricted specific epitopes recognized by cytotoxic T cells are peptides of defined sequences of amino acids and can be characterized with artificial intelligence and pattern recognition methods in combination with additional filters and optimization steps described herein. The predicted epitope peptides can be verified with biological assays for tumor or virus antigen-specific T cell activities using peripheral white blood cells of patients as source for the specific T cells. A composition of HLA-restricted specific antigenic peptides (1-100) for a particular virus or tumor together with adjuvants as CD4+ helper T cell activators can be used for effective vaccination.

A number of HLA-restricted tumor-specific epitopes and antigenic peptides for various cancers and viruses detected with the method of this invention is attached in the Tables.

Procedure:

a) Prediction of MHC-I specific epitopes

- Generation of a prediction tool for MHC-I binding and/or T-cell activation. This can be done by using any state of the art

technology for structure activity relationship (SAR) model generation, like ANN's, support vector machines (SVM's), SIMCA P, partial least squares projection to latent structures (PLS) etc.. As the basis for the application of these technologies a representative data set of peptides is used. This dataset, e.g., consists of peptides, known to bind to a given MHC-I molecule, e.g. those stored within the SYFPEITH database (Hans-Georg Rammensee, Jutta Bachmann, Niels Nikolaus Emmerich, Oskar Alexander Bachor, Stefan Stevanovic: SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics (1999) 50: 213-219) and peptides, that do not bind. Due to the fact, that there is only limited data on experimentally proven not-binding peptides a set of randomly generated peptides can be used for model generation, e.g. all epitopes, that can be generated out of the p53 protein. In this particular case ANN's were trained for HLA-0201; HLA-0101; HLA-1101, based on the epitopes given in SYFPEITH database using an evolutionary algorithm for optimization of weights and biases within the neural network. The criteria for using a generated SAR model for epitope prediction is the prediction quality of said model on a test dataset, that has not been used for training. The neural networks used within the next steps of this inventions were able to correctly assign almost all test data to the corresponding class (binding, not-binding).

Selection of the relevant antigenic proteins for various cancers and viruses.

This is done according to current state of the art technology and knowledge. The following criteria can be used for selection:

Proteins, described in literature as source of tumor associated antigens

Proteins, involved in apoptotic processes, e.g. p53

Proteins, belonging to tumor testis antigens and embryonic antigens, e.g. MAGE, BAGE, GAGE, CEA, AFP

Proteins, that are expressed in specific tissues, e.g. tyrosinase

- A procedure defining the degree of conservation for each potential epitope within the protein sequence, in particular, a procedure selecting (phylogenetically) conserved regions within a protein sequence.

This procedure consists of 3 steps:

1. Performing a similarity search against protein and/or nucleic acid data bases containing human and/or non-human sequences, e.g. by using BLAST (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402) FASTA or any other available tool.

See example in figure 1

2. Defining a similarity cutoff, e.g. when using BLASTP the "expect threshold" can be set to 1e-30. Only those proteins with a similarity higher then the selected cutoff are used to perform step 3.
3. Calculating the degree of conservation for each potential epitope. For this, the complete sequence of the selected tumor antigen is chopped into overlapping 9-mers (8-mers, 10-mers). For each of these epitopes a conservation score is calculated. This can be done by simply summing up the number of identical AA between the selected antigenic protein and the identified homologue proteins over all epitope positions. Alternatively substitution matrices, e.g. BLOSUM, PAM etc. (see. Altschul et.al.) can be used.

An example is given in figure 2.

- A procedure generating all possible peptide variants out of each epitope within the selected tumor antigen, by exchanging the natural amino acid at certain anchor residues by more preferred amino acids. In particular, an optimization step where amino acids (AA) within the so-called anchor residues of the MHC-I binding peptides are being exchanged. This procedure consists of 3 steps:
  1. Based on the knowledge about known epitopes (Hans-Georg Rammensee, Jutta Bachmann, Niels Nikolaus Emmerich, Oskar Alexander Bachor, Stefan Stevanovic: SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics (1999) 50: 213-219) or by using the so called "virtual alanine scan" technology (see PCT/EP01/14808) or by using any other technology the so-called "anchor residues" are identified. These are the positions within the epitope, that are most important for binding to the given MHC receptor.
  2. Moreover, by applying the same technologies, those AA, that are most preferable in these anchor positions are identified, e.g. for HLA-0201 the anchor position are position 2 and 9 with L , M, V and I (isoleucine) most preferred in the corresponding positions (according to Rammensee et al.). These preferred AA can also belong to the group of non-natural AA.
  3. The last step comprises the *in silico* generation of all possible peptide variants, e.g. for each epitope there are 8 peptide variants in case of 2 anchor residues with 2 different preferred amino acids each. These peptides are only virtually generated, so no peptide synthesis has to be applied at this stage of the process. When including non-natural AA so called peptidomimetics are generated.
- Evaluation of all potential epitopes generated within the previous steps by the SAR model, e.g. ANN's trained in step 1. In particular,

prediction of CD8+ T-cell epitopes, e.g. within an ANN. According to the results of the prediction the epitopes are ranked.

- The selection (filtering) of epitopes out of the ranked list is preferably done according to the following criteria:
  1. SAR model predict high MHC-I binding for the epitope, preferably the highest.
  2. The epitope is predicted to bind to more then one MHC-I molecule.
  3. The epitope has high conservation score, preferably the highest among all epitopes of a given tumor antigen.
  4. The epitope has the following properties:
     a. The epitope do not contain any of the following amino acids: P, M, C, G.
     b. The epitope does not contain four of the aliphatic amino acids (I; L;) in line, e.g ILLL is filtered out, but ILLFL is permitted.
     c. The epitope do not contain the sequences PEST in a line.

b) Verification of the predicted epitope peptides

- Verification of the predicted epitope with synthetic peptides and assays for the cytolytic activity and anti-tumor or anti-virus efficacy of the epitope-specific T cells using peripheral white blood cells of patients as source of specific T cells.

Those epitope selected according to part a) of the procedure are synthesized with standard procedures and tested in an in vitro assay, e.g. as described in PCT/DE99/00175 and Kern F. et al. Nature Medicine. (1998) 4(8):975-8,T-cell epitope mapping by flow cytometry. Those epitopes, that cause a specific T cell reaction within this assay are further developed into step c).

c) Vaccination with predicted epitopes

- Generation of vaccines that consist of 1-100, preferably 2-90, more preferably 5-80 and most preferably 10-50 relevant peptides as identified by a) and/or b) and optionally specific recall antigens as adjuvants for CD4* T cell stimulation and for induction of co-stimulation for the peptide and disease-specific CD8* cytotoxic T cells (CTL) or with adjuvants, co-factors or general CD4* T-cell stimulation antigens for co-stimulation of CD8* CTLs.
- In principle the epitopes identified within step a and b can be used in several vaccination strategies and are as such not restricted to the one mentioned above.
- Vaccination, in particular, intracutaneous or parenteral vaccination in humans with the vaccine pool.

There are two patents claiming the application of ANN for the prediction of MHC binding motifs of biologically active peptides and peptide mimetics (DE 198 26 442, WO 98/53407 C2).
The method presented within this invention preferably combines the application of ANN with two additional steps:

- An optimization step where amino acids (AA) within the so-called anchor residues of the MHC-I binding peptides are being exchanged
- A procedure selecting conserved regions within a protein sequence.

The optimization and the selection procedure can apply knowledge and/or computer-based algorithms.

This invention provides the following advantages in comparison to previously described methods for T-cell epitope prediction:

- The epitopes yielding highest CTL response in most human individuals will be the least variable ones and therefore be of the highest pharmacological relevance.
- The specific optimization step will improve the MHC-binding properties of the peptides without affecting the biological activity of the peptide. The application of this optimization procedure to all 9-mers (8-mers, 10-mers) of a given tumor antigen allow the identification of previously not identified epitopes and mimitopes. Further, it is possible to obtain biologically active peptides that differ from naturally occurring sequences.
- The parallel prediction of binding to several different MHC-I molecules allows the identification of epitopes, that have a significant higher application potential.
- The application of knowledge based filters (PEST sequences; non tolerated amino acids etc.) increase the probability of biological effects and application potential.
- The usage of *in vitro* assays for the verification of the epitopes that, based on the biological reactivity of cytotoxic T cells of cancer or virus infection patients, ensures detection of disease-relevant specificities
- The usage of state of the art pattern recognition technologies in combination with the afore mentioned steps yield in a higher prediction accuracy.
- For vaccination, 1-100 peptides related to a particular virus or cancer, will be used as a vaccine. Additionally, specific co-factors, adjuvants and CD4+ T-cell antigen for co-stimulation of CD8+ T-cells will be included. This can be applied intracutaneously, parenterally, etc.

Fig.1 schematically shows a similarity search, and

Fig.2 shows an example of calculation of conservation scores.

Examples

Example 1

The performance of the method of the invention will be explained in the following by way of an example.

First, an antigenic protein is selected, e.g. from a database. In the case of this example, a protein having 509 amino acids is chosen as an antigenic protein. Said protein is is fragmented virtually (by computer) to give 500 peptides having a length of 9 amino acids each. A conservation score is determined for each of these 9-mers. In the subsequent optional step anchor positions and preferred amino acids at these positions are determined. In the case of this example it is assumed that anchor positions are at positions 2 and 9 and 2 optimal amino acids each are described in the prior art for each position. This leads to 8 variants for each 9-mer, so a total of 4,500 epitopes are present (8 variants and 1 original). These epitopes are now tested as to whether they are CD8+ T-cell epitopes by means of a pattern recognition technology, e.g. SAR and ANN, respectively. In particular, MHC binding capacity can be determined this way.

Assuming it is found that 300 epitopes are effective, the conservation score of these 300 epitopes is now used to determine the best 100 epitopes.

Subsequently, a filter can be used which sorts out particular peptides, e.g. peptides containing proline (because of unfavorable folding) and peptides, in the case of which synthesis problems are to be expected.

In this way the number of epitopes can be further reduced, e.g. to 50. These 50 epitopes can now be verified in an in vitro assay for their

activity. Part or all of the peptides verified as being active can then be pooled and used as a vaccine.

Example 2

In vitro verification of the T-cell activation functionality of peptides identified or optimized, respectively, according to the invention.

Peptide sequence	Source protein	Frequencies reactive CD8+ T cells	
		Melanoma	Cutaneous T-cell lymphoma
VTAQVVLQA	GP100	0,08	0,04
VLAQVVLQL	GP100 optimized	0,18	0,12
LVHFLLLKY	MAGE	0,99	0,03
LLHFLLLKL	MAGE optimized	1,10	0,03
FVWLHYYSV	TYR2	1,01	0,01
FLWLHYYSL	TYR2 optimized	0,82	0,02
Control		0,10	0,02

Claims

1. A method for providing, identifying or/and optimizing peptides which induce cytotoxic T-lymphocytes, comprising the steps:
   (a) selecting one or more antigenic proteins,
   (b) selecting conserved regions within the protein sequence of the one or more antigenic proteins, and
   (c) identifying CD8+ T-cell epitopes within the protein sequence of the one or more antigenic proteins.

2. The method according to claim 1, further comprising the step:
   (d) optimizing the identified CD8+ T-cell epitopes by exchanging one or more amino acids at the anchor positions thereof.

3. The method according to claim 2, wherein step (d) is performed prior to step (c).

4. The method according to any of the preceding claims, wherein step (c) is performed using an artificial neural network.

5. The method according to any of the preceding claims, wherein in step (a) one or more antigenic proteins for cancer or/and a virus are selected.

6. The method according to any of claims 1 to 5, wherein peptides having 4 to 30 amino acids are obtained.

7. The method according to any of the preceding claims, wherein an additional filtering step is applied.

8. The method according to any of claims 1 to 7, further comprising the step:
   - verification of the activity of the identified or/and optimized peptides in vitro.

9. Pharmaceutical composition comprising one or more peptides which induce cytotoxic T-lymphocytes obtainable according to the method of any of claims 1 to 8.

10. The pharmaceutical composition according to claim 9, further comprising adjuvants, co-factors and/or co-stimulating agents.

11. A method for providing a pharmaceutical composition for the induction of cytotoxic T-lymphocytes, comprising:
   (a) providing one or more peptides which induce cytotoxic T-lymphocytes according to the method of any of claims 1 to 8, and
   (b) using the one or more peptides for the manufacture of a pharmaceutical composition.

12. Isolated peptide as depicted in any of Tables 1, 2, 3 or 4, including the variants generated by AA exchange at positions 2, 6 and/or 9.

13. Isolated peptide having the formula VTAQVVLQA, VLAQVVLQL, LVHFLLLKY, LLHFLLLKL, FVWLHYYSV or FLWLHYYSL, including the variants generated by AA exchange at positions 2, 6 and/or 9.

14. Pharmaceutical composition comprising one or more peptides according to claim 12 or 13.

15. Use of a peptide according to claims 12 or 13 or obtainable according to the method of any of claims 1 to 8 for the manufacture of a pharmaceutical for the induction of cytotoxic T-lymphocytes.

16. Use according to claim 15 for the prevention, treatment or diagnosis of cancer or viral infections.

Figure 1:

Similarity search with selected tumor associated antigen using BLASTP against SWISS-PROT

```
BLASTP 2.1.3
Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query=    (385 letters)

Database: Non-redundant SwissProt sequences
          96,469 sequences; 35,174,128 total letters

                                                               Score    E
Sequences producing significant alignments:                    (bits)  Value

CD34_HUMAN  HEMATOPOIETIC PROGENITOR CE...  543  e-154
CD34_CANFA  HEMATOPOIETIC PROGENITOR CE...  359  8e-99
CD34_MOUSE  HEMATOPOIETIC PROGENITOR CE...  349  9e-96

Alignments
1        17  WTALCLLSLLPSGFMSLDNNGTATPELPTQGTFSNVSTNVSYQETTTPSTLGSTSLHPVS 76
3183511  17  ............................................TTTPSTLGSTSL.... 76
2498215  17  ...........F..TNTETVI.P.TVPTSTEIM.A..E.T.KR.AITLTPSGTTTLYS.. 76
3182946  17  .V....M..-----H.N.LTS..T.TS...ISPS.P..E.VE.NITSSIPGSTSHYLIY 71

1        77  QHGNEATTNITETTVKFTSTSVITSVYGNTNSSVQSQTSVISTVFTTPANVSTPETTLKP 136
3183511  77  ............................NTNSSVQSQTSVISTVFTT............. 136
2498215  77  .DSSGT.AT.S....HV....E..LTP.TMNSSVQSQTSLAITVSFT.T.F..SSV..E. 136
3182946  72  .DSSKT.PA.S..M.N..V..G.P.GS.TPHTFSQPQTSPTGILPTTSDSI..S.M.W.S 131

1       137  SLSPGN------V--SDLSTTSTSLATSPTKPYTSSSPILSDIKAEIKCSGIREVKLTQG 188
3183511 137  ......------.--SDLSTTSTSLATSPTKPYTSSSP...................... 188
2498215 137  ..L...GSDPPYN--STSLVTSPTEYYTSLSPTPSRNDTP.T..G......VK....N.. 194
3182946 132  ..PSI.------.SDYSPNNSSFEMTSPTEPYAYTSSSAP.A..G..........R.A.. 185

1       189  ICLEQNKTSSCAEFKKDRGEGLARVLCGEEQADADAGAQVCSLLLAQSEVRPQCLLLVLA 248
3183511 189  ............................................................ 248
2498215 195  ....L.E....ED....NE.K.TQ...--.KEP.E...G.............H....... 252
3182946 186  ....LSEA...E....EK..D.IQI..EK.E.E.....S.............E...M... 245

1       249  NRTEISSKLQLMKKHQSDLKKLGILDFTEQDVASHQSYSQKTLIALVTSGALLAVLGITG 308
3183511 249  ............................................................ 308
2498215 253  .K..LF.....LR..........R.......G......R...........I......T.. 312
3182946 246  .S..LP......E......R....QS.NK..IG......R..........V...I..T.. 305

1       309  YFLMNRRSWSPTGERLGEDPYYTENGGGQGYSSGPGTSPEAQGKASVNRGAQENGTGQAT 368
3183511 309  ........................GGGQGYSSGPGTS....................... 368
2498215 313  ........................GGGQGYSSGPGVS............P......... 372
3182946 306  ........................GGGQGYSSGPGAS..T....N.T............. 365

1       369  SRNGHSARQHVVADTEL 385
3183511 369  ................. 385
2498215 373  ..........M...... 389
3182946 366  ................. 382


  Database: Non-redundant SwissProt sequences
    Posted date:  May 11, 2001  5:54 AM
  Number of letters in database: 35,174,128
  Number of sequences in database:  96,469

Lambda     K      H
```

```
  0.312     0.128     0.357

Gapped
Lambda     K      H
  0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 19900574
Number of Sequences: 96469
Number of extensions: 647916
Number of successful extensions: 1005
Number of sequences better than 10.0: 4
Number of HSP's better than 10.0 without gapping: 3
Number of HSP's successfully gapped in prelim test: 1
Number of HSP's that attempted gapping in prelim test: 998
Number of HSP's gapped (non-prelim): 4
length of query: 385
length of database: 35,174,128
effective HSP length: 56
effective length of query: 329
effective length of database: 29,771,864
effective search space: 9794943256
effective search space used: 9794943256
T: 11
A: 40
X1: 16 ( 7.2 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 42 (21.9 bits)
S2: 66 (30.0 bits)
```

Figure 2:

Calculation of 2 different conservation scores for all possible epitopes within position 25 – 78 of the query molecule CD34_HUMAN, when using BLASTP as shown in figure 1.

```
CD34_HUMAN  HEMATOPOIETIC PROGENITOR CE...  543  e-154
CD34_CANFA  HEMATOPOIETIC PROGENITOR CE...  359  8e-99
CD34_MOUSE  HEMATOPOIETIC PROGENITOR CE...  349  9e-96
```

Pos.	Sequences	# Identities to query	# identities and second most frequent AA	Conservation Score 1	Conservation Score 2
25	L...	4	4	34	34
26	L...	4	4	34	34
27	P..-	3	3	34	34
28	S.F-	2	2	32	32
29	G..-	3	3	31	31
30	F..-	3	3	30	30
31	M.T-	2	2	28	28
32	S.NH	2	2	27	27
33	L.T.	3	3	26	26
34	D.EN	2	2	24	24
35	N.T.	3	3	23	23
36	N.VL	2	2	22	22
37	G.IT	2	2	22	22
38	T..S	3	3	22	22
39	A.P.	3	3	22	22
40	T...	4	4	24	24
41	P.TT	2	4	24	26
42	E.V.	3	3	24	26
43	L.PT	2	2	24	26
44	P.TS	2	2	23	25
45	T.S.	3	3	24	26
46	Q.T.	3	3	25	27
47	G.E.	3	3	25	27
48	T.II	2	4	24	28
49	F.MS	2	2	22	26
50	S..P	3	3	23	25
51	N.AS	2	2	22	24
52	V...	4	4	24	26
53	S..P	3	3	25	27
54	T.E.	3	3	25	27
55	N...	4	4	26	28
56	V.TE	2	2	25	27
57	S...	4	4	27	27
58	Y.KV	2	2	27	27
59	Q.RE	2	2	26	26
60	E...	4	4	28	28
61	TTAN	2	2	26	26
62	TTII	2	2	25	25
63	TTTT	4	4	26	26
64	PPLS	2	2	24	24
65	SSTS	3	3	25	25
66	TTPI	2	2	23	23
67	LLSP	2	2	23	23
68	GGGG	4	4	25	25
69	SSTS	3	3	24	24
70	TTTT	4	4	26	26
71	SSTS	3	3	27	27
72	LLLH	3	3	26	26
73	H.YY	2	4	26	28
74	P.SL	2	2	25	27
75	V..I	3	3	26	28
76	S..Y	3	3	27	29
77	Q...	4	4	27	29
78	H.DD	2	4	26	30